



**CENTRE
DE RECHERCHE
SUR LES TRANSPORTS**

(CRT)

**CENTRE
FOR RESEARCH
ON TRANSPORTATION**

UNIVERSITÉ DE MONTRÉAL
ÉCOLE DES HAUTES ÉTUDES COMMERCIALES
ÉCOLE POLYTECHNIQUE



C.P. 6128, SUCCURSALE A
MONTRÉAL (QUÉBEC)
CANADA
H3C 3J7

TÉLÉPHONE : (514) 343-7575
TÉLEX : 05-24146 BIBPOLYTEC
TÉLÉCOPIEUR (FAX) : (514) 343-7121

CRT-657

**A GLOBAL OPTIMIZATION ROUTING MODEL AND
APPLICATIONS IN TELECOMMUNICATION NETWORKS**

by

**Brunilde Sanso
and
François Soumis**

**Université de Montréal - HEC - POLY
Centre de recherche sur les transports - Publication #657
October 1989**

90-01616/

Dépôt légal - 4e trimestre 1989

ABSTRACT

We present a centralized optimization model for the routing of calls in circuit switched networks. The problem can be stated as a multicommodity model without capacity constraints that takes into account the stochastic character of the demand. This model is solved using a Frank-Wolfe algorithm. Several applications of the model are outlined in the paper. Results on routing policy evaluation and network reliability are shown.

Keywords: routing, telecommunication, circuit switching, multicommodity problem, performance, reliability.

- - - - -

RÉSUMÉ

Nous présentons un modèle centralisé de routage d'appels pour des réseaux de commutation par circuit. Il s'agit d'un problème de multiflot sans contrainte de capacité qui tient en considération le caractère stochastique de la demande. Le problème est résolu avec un algorithme de Frank-Wolfe. Le papier montre des applications possibles du modèle et de plusieurs résultats en évaluation du routage et en fiabilité du réseau.

Mots clés: acheminement d'appels, routage, multiflot, performance, fiabilité, télécommunication, commutation par circuit.

1. INTRODUCTION

In telecommunication networks the routing is the procedure which determines how a call is routed. This procedure can widely affect network performance in terms of costs and reliability. This paper focuses on the routing models, the algorithm used to evaluate the outcome of the routing and on how that outcome can affect network design and reliability.

Routing procedures are composed of three stages (see [4]). The first involving the identification of paths available for routing a particular call. The second involving the selection of a path to be employed by the call. This requires rules specifying what happens to a call upon arriving in the network. Packet switched networks are characterized by an extra stage at this point in which it is determined if a call should be routed at all even if a path is available, whereas in circuit switched networks it is supposed that if a path is available the call will be routed. The third stage of the routing procedure involves the management of blocked calls.

It is as a function of its second stage that a routing procedure is classified. A popular class of routing models employs fixed alternate routing. Under this rule attempts are made to route a blocked call via a second pre-established route, via a third, and so on, until it is lost. Another class of models employees the routing without overflow and with load share rule. Under this rule, demand for every origin-destination pair is assigned to a set of paths according to a coefficient characteristic of the paths; calls that cannot be assigned to a path are lost.

Fixed alternate routing constitutes a simple rule for handling calls upon their arrival. It is, however, a rigid scheme in which there is no opportunity for intervention with respect to performing routing optimization. The load share rule, on the other hand, permits this intervention

allowing traffic assignement to be directed by varying attraction coefficients of routes, for exemple. Fixed alternate routing is in general use in circuit switched networks today, but load share routing is becoming increasingly popular due to its flexibility.

The routing classification described above is a function of telecommunication technology. The model we will use to evaluate system performance, however, corresponds to a global optimization procedure which is independent of technology. In it, the three stages of routing a call are incorporated into one and our goal is to optimize the global performance of the system. It is important to point out that this type of model is an ideal one, the routing, under this model, cannot be actually implemented in a real network. However, this type of model allow the evaluation of network performance under ideal operation conditions. Therefore, the study of network performance under this kind of model can give an idea on wether it is worth it or not to develop more flexibly routing technology that could be closer to global optimization.

2. THE MODEL

Several deterministic global optimization models have been presented in the past (see [2], [3], [4]). The term deterministic describes the assumption made with respect to demand in these models. They consist of a linear multicommodity problem with explicit capacity constraints. The objective is the minimization of lost demand over all commodities, and the flow in the network is the traffic per commodity per arc. Not only are these problems difficult to solve but these models are not reflective of reality due to their assumption of deterministic demand.

The global optimization model for testing network performance, proposed here, is of stochastic nature. Capacity constraints have been replaced by a penalty function in order to simplify the model and, at the same time, reflect the stochastic character of the demand.

Let

N : node set for the entire graph;

M : arc set for the entire graph;

n : cardinality of N ;

m : cardinality of M ;

O : origin node set;

D : destination node set;

D_p : set of destination nodes for all flow originating at node p ;

d^{pq} : demand of origin p destined to node q ;

$\Gamma(i)$: set of arcs originating at node i ;

$\Gamma^-(i)$: set of arcs terminating at node i ;

C_u : capacity of arc u ;

x_u^p : traffic flow of origin p passing through arc u .

The global optimization model is thus formulated:

$$\min Z = \sum_u P_u(X) \quad (1)$$

s.t.

Flow conservation

For all $p \in O$, for all $q \in D$, for all $i \in N$

$$\sum_{u \in \Gamma(i)} x_u^p - \sum_{u \in \Gamma^-(i)} x_u^p = \begin{cases} d^{pq} & \text{if } i = p \\ -d^{pq} & \text{if } i = q \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

Nonnegativity

$$x_u^p > 0 \text{ for all } u \in M, \text{ for all } p \in O \quad (3)$$

$P_u(X)$ of (1) above is the penalty function which replaces the capacity constraints to yield a stochastic global optimization model. In fact, we take the penalty function to be the lost call traffic function since our objective of evaluating network performance involves its minimization in any case. $P_u(X)$, Erlang's lost call traffic function, can be expressed as a function of the blocking probability of the system:

$$P_u(X) = \sum_{p \in O} x_u^p B(C_u, \sum_{p \in O} x_u^p) \quad (4)$$

To simplify notation, we allow:

$$X_u = \sum_{p \in 0} X_u^p \quad (5)$$

such that the B component of (6) can be expressed as:

$$B(C_u, X_u) = \frac{X_u^{C_u} / C_u !}{\sum_{i=0}^{C_u} X_u^i / i !} \quad (6)$$

$B(C_u, X_u)$ corresponds to the probability that a call will be blocked at a link having C_u trunks of capacity and where X_u is the offered traffic.

The objective of minimizing the lost call traffic (LCT) involves the minimization of:

$$Z = \frac{\sum_u X_u^{C_u+1} / C_u !}{\sum_{i=0}^{C_u} X_u^i / i !} \quad (7)$$

This function has been proven to be convex over the variable X_u (see [3] and [5]). Z is a function that can be separated by arc. This facilitates solution of the problem defined by (1) to (3) but introduces error in the model. The probability of losing a unit of flow on path $u_1, u_2, \dots, u_k$ is implicitly considered to be:

$$\sum_{j=1}^k B(C_{u_j}, \sum_{p \in 0} X_{u_j}^p) \quad (8)$$

This approximates reality only if blocking probabilities are very low because the use of a function additive by arc requires that the probability of blocking on two or more arcs of the same path be negligible.

A more correct estimate of the probability of blocking on a path would be:

$$[1 - \prod_{j=1}^k (1 - B(c_{u_j}, \sum_{p \in O} x_{u_j}^p))] \quad (9)$$

Unfortunately, expression (9) is not only not separable by arc but, is erroneous as well since it derives from the assumption that the flows on the arcs are independent variables.

Modifications to the routing model

An advantage of using a global optimization model is that one has available the possibility of routing calls in the manner that one sees as the most convenient. For instance, it might prove worthwhile to avoid paths that are too long in routing calls. Alternatively, attempting to equilibrate the lost call traffic per origin might prove beneficial.

In order to avoid lengthy paths in routing calls a simple shortest paths term can be added to the objective function (7). The second alternative is more difficult to incorporate since the objective function is separable by arc and would thus be impossible to introduce a term related to the variance of the lost call traffic over all network commodities. One way in which this alternative can be handled, however, is to add a term such as (T_1) to the objective function:

$$T_1 = \sum_u p_u^2(x) \quad (10)$$

Unfortunately, the use of (10) might include the uniformization of the LCT by removing flow from high capacity arcs and blocking the flow on low capacity arcs. This situation can be rectified by using, instead, a term accounting for arc capacities:

$$T_2 = \sum_u (P_u(x) / C_u)^2 \quad (11)$$

Summarizing, the objective function of the global optimization model to be used here is:

$$Z = \sum_u (K L_u(x) + P_u(x) + M(P_u(x) / C_u)^2) \quad (12)$$

Where L_u is a shortest path term, and K and M are arbitrary weights given to the shortest path term and the equalization terms respectively.

Applications of the routing model

The first application of the model presented is in studying the outcome of the routing itself, that is, the LCT, the arc flows and the routes taken by every commodity when the global optimization routing policy is put to work.

Another application of the model is in studying network performance which take us to the field of Network Reliability. In [2] a new measure of performance, the expected lost call traffic (ELCT) was introduced and a methodology to evaluate it was presented.

The idea was to consider a failure scenario and, for each scenario, solve the routing problem and evaluate its LCT. If the scenario were chosen such that they represented the most probable states of the system, bounds on the ELCT were easily computed (see [2], [3], [6]). Enumeration

procedures to evaluate those states have been exposed for both binary failure [6] and multiple failure component systems [7]. It has been shown, however, that since telecommunication networks have high component reliability, no failure and single failure states cover most of the state space (see [2], [3]). Therefore, no special enumeration procedure is necessary to evaluate the most probable states of the system.

Another important application of the routing model is in Network Design and Dimensionning. In fact, it is evident that reliability and economic factors can be widely affected by the routing policy. Figure 1 summarizes the interaction between the routing policy, reliability and design.

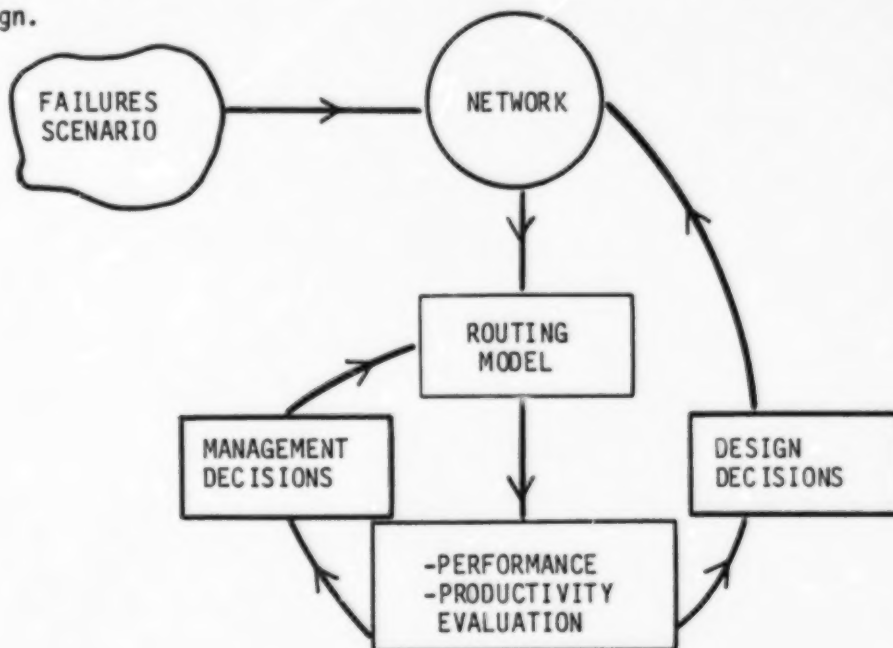


Figure 1-

In this paper we will focus on the routing and the reliability evaluation of an existing network. The design application will be studied in a forthcoming publication. The general idea being that for a given routing policy and level of performance, those arcs that do not respect that level will be redimensioned.

3. THE ALGORITHM

The algorithm use to solve the routing model is a Frank-Wolfe. The implementation is similar to the one used in traffic equilibrium assignment problems in which the variables kept at every step of the procedure are the total flows on the arcs (see, for instance, [8]).

The convergence of the algorithm is there guarantee if the objective function is convex over the variable representing the total flow on the arcs. In other words in order to prove the convexity of (12), $P_u(X)$ had to be proven convex over the variable X_u . Two different proofs are presented in [3] and [5].

Approximation of the objective function

The complexity of $P_u(X)$ (expression (7)) was such that it was convenient to make an approximation of the function in order to reduce calculation time for large scale problems.

We remind the reader that $P_u(X)$ is a two variable function, the arc capacity C_u and the arc flow X_u . Since our routing problem consists of determining the traffic when arc capacities are fixed, we have considered P_u as a one variable function. Nonetheless, it is necessary to approximate $P_u(X)$ for every possible arc capacity value.

The first problem encountered in approximating $P_u(X)$ was to actually obtain the real values of the function for high levels of capacity (the direct evaluation of (7) may need computing so large expressions such as 10 000 !). Thus, the evaluation of $P_u(X)$ was done by using a computational trick similar to the one shown in [9] to compute Erlang's B formula.

To ease the notation, let x represent the total arc flow on a specific arc of capacity C , $P(x)$ the LCT on that arc and $F(x)$ the inverse of $P(x)$, then $F(x)$ can be expressed as:

$$F(x) = \frac{1}{P(x)} = \sum_{t=0}^C \frac{C!}{(C-t)!} \frac{1}{x^{t+1}} \quad (13)$$

$F(x)$ can be reformulated as:

$$F(x) = \sum_{j=0}^C A_j \quad (14)$$

Where

$$A_j = \begin{cases} \frac{1}{x} & \text{if } j=0 \\ \frac{[C-(j-1)]}{x} A_{j-1} & \text{if } j=1, \dots, C \end{cases} \quad (15)$$

Expression (15) allow the evaluation of $P(x)$ by means of a loop performing a simple multiplication C times.

$P(x)$ was found to behave as a three piece function: almost zero at the first interval, quadratic at the second and linear at the third. A simple theoretical calculation shows that the function tends toward its asymptotes ($P(x) = 0$ if $x < C$ and $P(x) = x - C$ if $x > C$) if C becomes very large. $P(x)$ was thus approximated as:

$$P(x) = \begin{cases} 0 & \text{if } x < C - \epsilon \\ A x^2 + B x + \zeta & \text{if } C - \epsilon < x < C + \epsilon \\ x - C & \text{if } x > C + \epsilon \end{cases} \quad (16)$$

In order for $P'(x)$ to be continuous it was found that

$$A = \frac{1}{4\epsilon} ; B = \frac{\epsilon - C}{2\epsilon} ; \zeta = \frac{(C - \epsilon)^2}{4\epsilon} \quad (17)$$

and ϵ was chosen so that it is made smaller as C increases. Figures 2 and 3 depict the goodness of the approximation for two values of C .

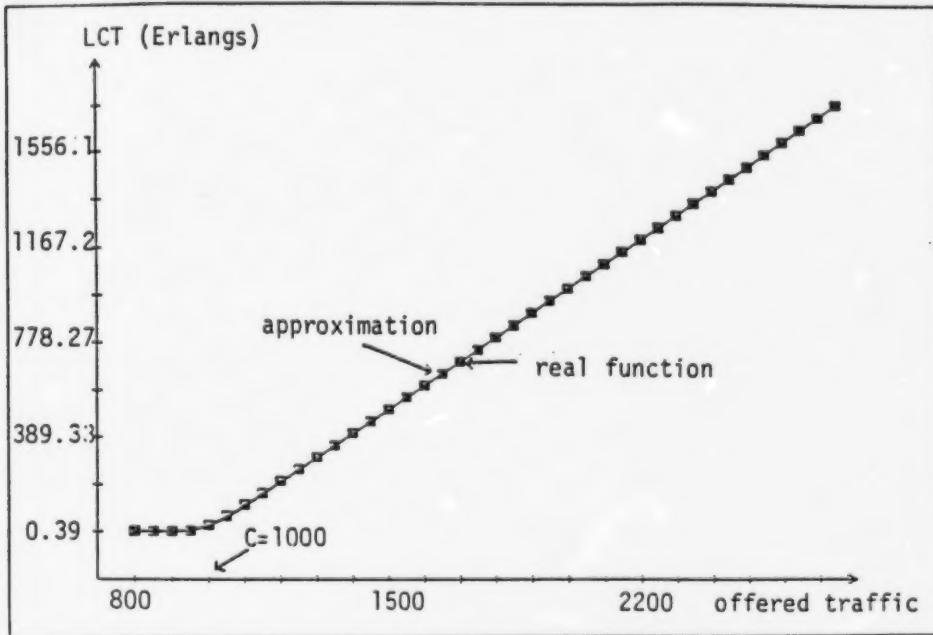


Figure 2 - Approximation of $P(x)$ for $C = 1000$

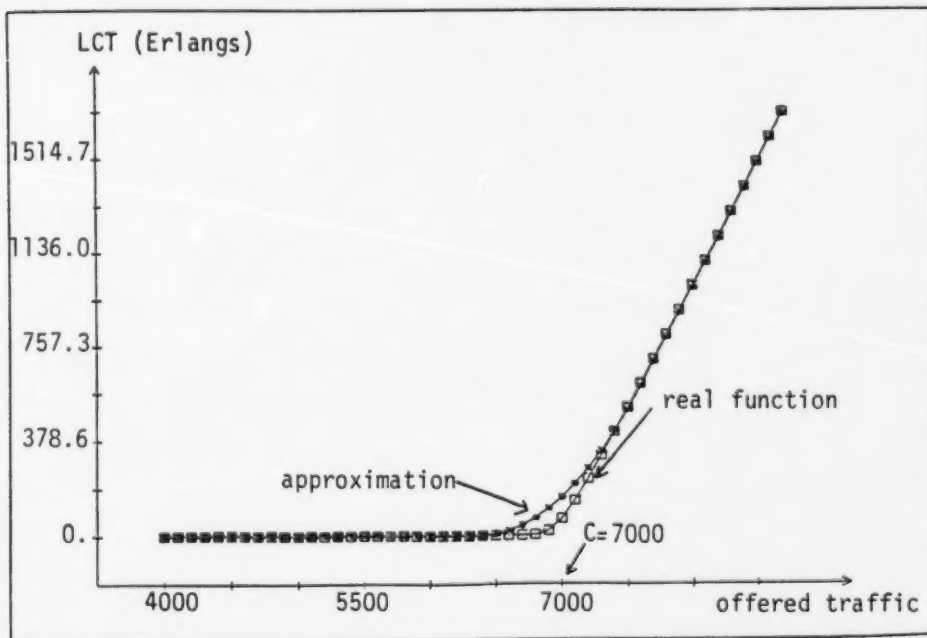


Figure 3 - Approximation of $P(x)$ for $C = 7000$

It is important to point out that the "bad" approximation made by the quadratic function when C is very large (see figure 3) was made on purpose. In fact we tried to avoid the "square" asymptotic behavior of $P(x)$ around the point $x=C$ for large values of C since it could cause problems at the optimization level.

4. RESULTS

In this section, firstly, the network will be presented, then results of the routing outcome and network performance will be shown.

The network

Figure 4 depicts the RED network which is a physic network constructed from the following data supplied by a telecommunication company:

Number of nodes	= 40
Number of unoriented arcs	= 62
Average connectivity	= 2 to 3
Cable lenght	= 5 Km
Arcs lenght	= 5 to 15 Km
Arcs capacity	= 1000 to 10000 circuits
Average demand between switching stations	= 50 circuits
Variance of the demand	= 10 to 1000 circuits
Mean rate of failures	= 2 failures per year per 1000 Kw of cable
Mean repair time	= 15 hours per failure.

In figure 4 the nodes represent the switching stations and are numbered on the corner of each square. Arc capacities (in thousand of circuits) are shown in brackets on every arc. Arc lenghts (in units of 5 Km) are represented by vertical dashes on the arcs.

Routing results

The results that can be obtained by using the routing package are shown in figure 5. First, two types of variations will be studied: one related to the level of demand and the second one related to the choice of the objective function. The outputs are arc flows, LCT, the objective function and the mean number of arcs needed to route a call.

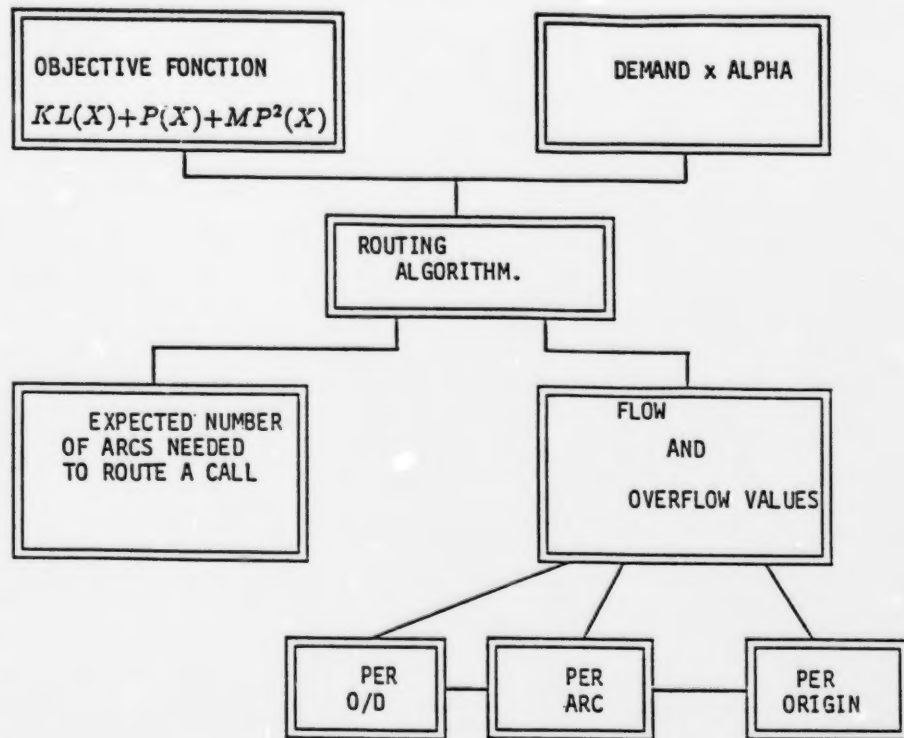


Figure 5 - Features of the Routing Package

The first results to be shown correspond to the variation of the CPU time needed to solve the routing problem over the level of demand. Figure 6 depicts two curves, one for the algorithm dealing with flows per origin-destination and one for the algorithm dealing with flows per origin. The behavior is exactly the same: for small and high levels of demand the CPU time is relatively short however, once the demand approaches 80% of the basic demand there is a high increase in CPU time. This behavior is

explained by the shape of the objective function and by the characteristics of the Frank-Wolfe algorithm. The objective function for one arc was approximated by two linear functions at the extremes (for small and big values of X) and by a quadratic function in an interval including the arc capacity. The shape of the global objective function must be similar to the one described for the arc function. Therefore when the demand is very low or very high the economic function must be almost linear and the Frank-Wolfe algorithm will provide results in very few iterations. On the other hand, when the demand is such that the algorithm has to work in the non-linear part of the function, many iterations are needed to attain the optimum.

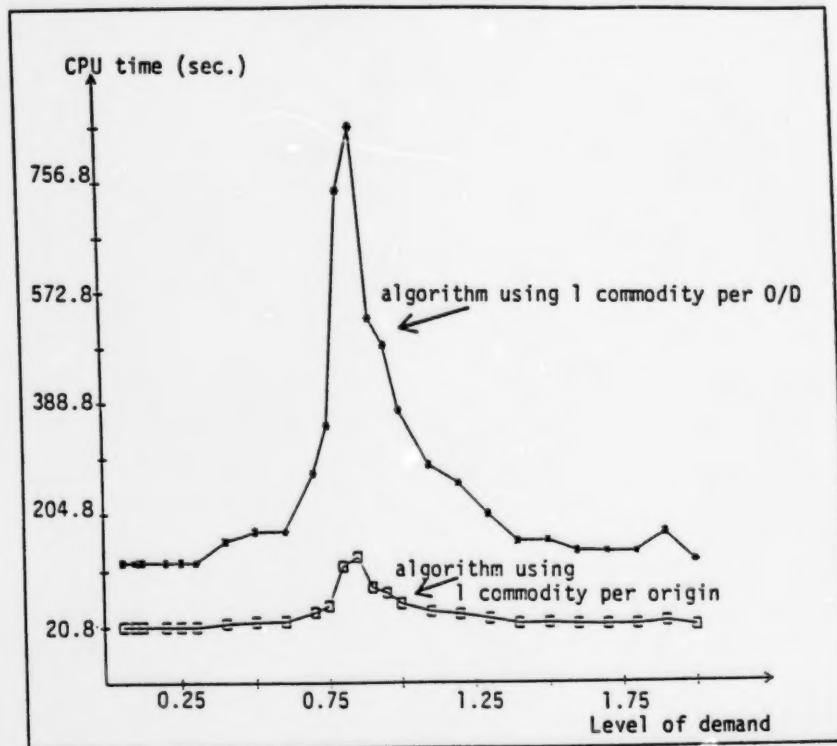


Figure 6 - CPU time VS level of demand

The second result to be studied is the mean number of arcs needed to route a call when the weight of the shortest path term (SPT) of the objective function is increased. The test was performed for three different levels of demand: 80%, 100% and 200% the basic demand. The results, depicted in figure 7 show that, as expected, the addition of a shortest path term to the economic function will have a definite impact on the mean number of arcs needed to route a call (MNARC). The most dramatic decrease in the MNARC is obtained when the demand is 80% the basic demand, on the other hand the lower benefic impact is gotten for 200% the basic demand. This can, again, be explained by the shape of the objective function for the different levels of demand. In fact, it has already be seen that when the demand approaches 80% of the basic one the objective function presents a definite non-linear character, hence, the introduction of a linear term such as the SPT will have a strong influence on the outcome of the routing. On the other hand, for very high levels of demand the objective function presents an almost totally linear behavior, the calls would be "naturally" assigned to the shortest paths and therefore the addition of a SPT will make a mild difference on the routing. Another interesting result to be pointed out is the asymptotic behavior of the MNARC over the weight given to the SPT. This indicates that there is a lower bound on the mean number of arcs needed to route a call which is inherent to the topological conditions of the graph and independent of the level of demand.

The third test to be performed on the routing was the influence of the quadratic term of the objective function over the variance of the percentage of demand lost on every arc of the network. We remind the reader that the original idea of the test was to equalize losses for all the

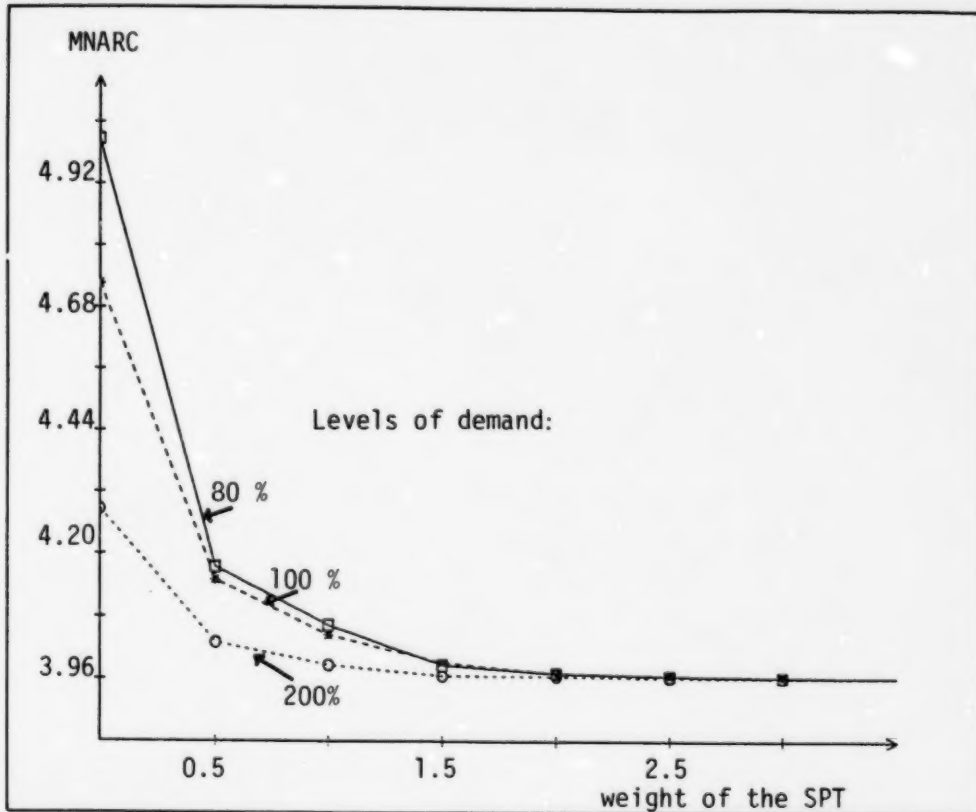


Figure 7 - Variation of MNARC over the weight of the SPT

commodities of the network. However, since the algorithm to be used worked at the arc level, and not the commodity level, the only egalization of the losses that could be attained was over the arcs of the network. In figure 8 the behavior of the variance of the % of demand lost on the arcs over M , the weight of the quadratic term in (12) is depicted. The % of demand on arc u being defined by $[P_u(X)/C_u] * 100$. As expected, the increase of M make decrease the variance of the lost demand over the arcs.

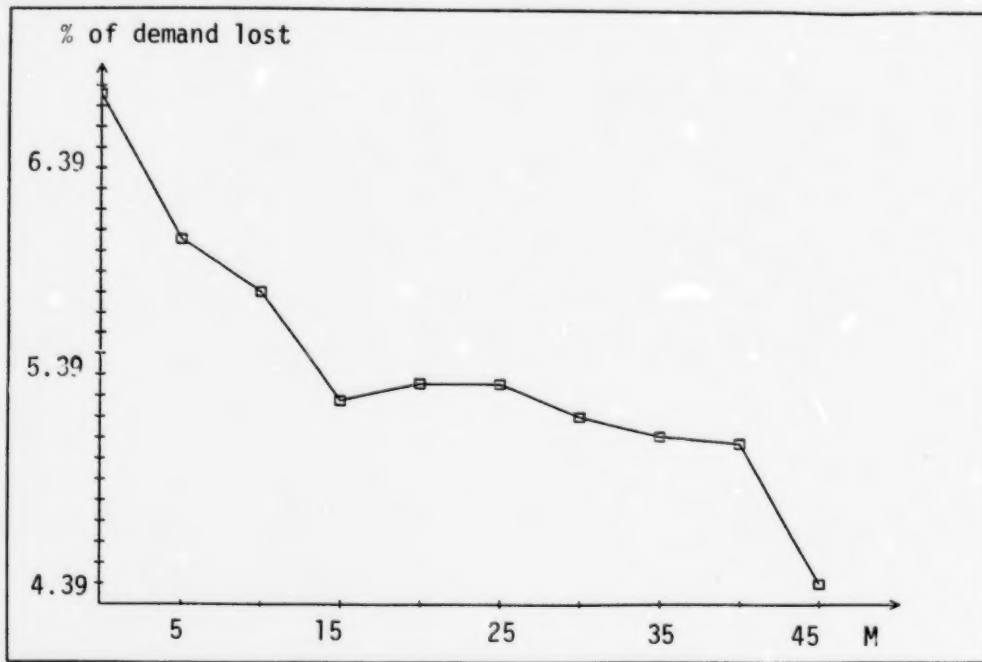


Figure 8 - Variance of the % of demand lost on the arcs vs M.

Reliability Results

In this section we will show two results that underline the importance of the routing in evaluating the measure of reliability. First, in figure 9, the upper and lower bounds on the expected lost call traffic are depicted over the failure time for different stopping criterias of the routing algorithm.

For every case two stopping criterias were chosen, one for the no-failure state and one for the single failure states, the idea being that failure probabilities are so low in telecommunication networks that too much precision at the 62 single failure states was not worth the amount of

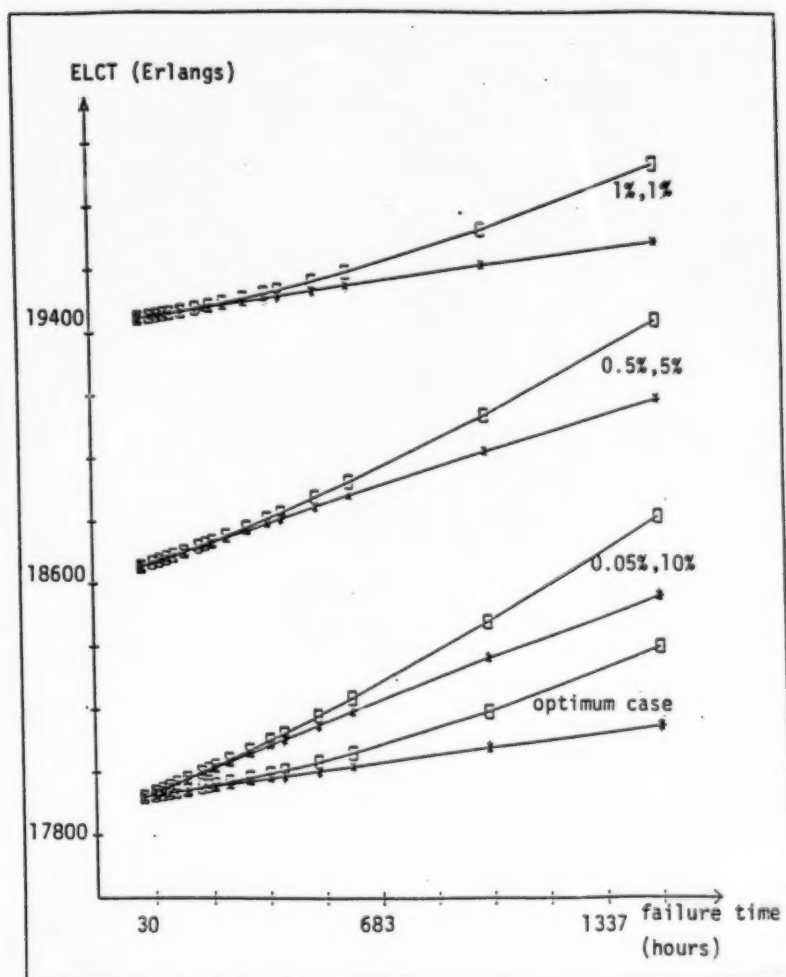


Figure 9 - Lost call traffic bounds vs failure time

calculation time. In fact it is shown that better results are obtained for the (0.05%, 10%) criteria case than for the (0.5%, 5%) case, the former being less time consuming (1353 seconds versus 1473.5 seconds of CPU on a SUN 3/50).

The second result that focuses on the importance of the routing is the comparison of the ELCT with and without optimization. In fact, when a

failure takes place on an arc, two different policies are possible: either re-route the demand that was offered to the arc or loose that demand at the switching point immediatly before the failed arc. When the network is, even in the no-failure state, very congested, the re-routing option could cause overloading in adjacents links and could be detrimental.

However if the network has been well designed and work at a low congestion level re-routing the calls after a failure is definite beneficial. This is clearly seen in figure 10 in which the upper bounds on the ELCT are shown for two routing policies: with and without re-optimization after a failure.

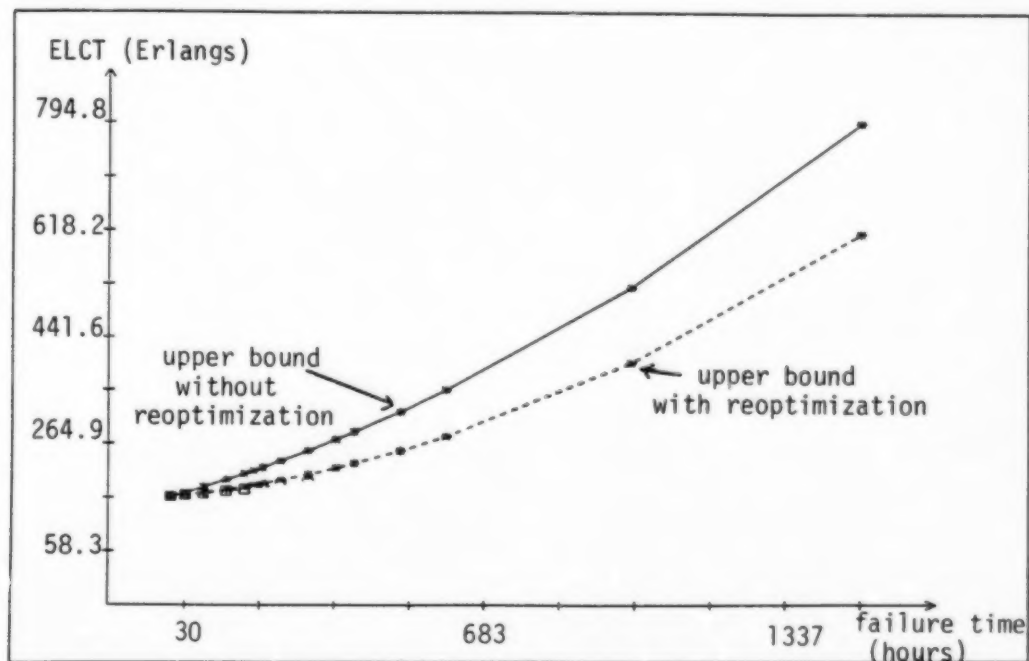


Figure 10 - Comparison of bounds with and without the re-optimization policy

5. CONCLUSION AND FURTHER WORK

We have presented a global optimization routing model which predicts network behavior and evaluate network performance. This model corresponds to an idealistic routing policy since actual technology do not allow a completely centralized routing. This type of model, however, is useful for the study of the utility of improving modern routing technology so that it would tend toward global optimization routing.

The importance of the routing policy when evaluating network performance has been underlined. In fact, we believe that reliability should no longer be considered as a "static" measure of network performance but rather should reflect the flexibility of the network to adjust to failure. In other words the reliability of the network is a function not only of network topology but of the routing policy as well.

Regarding further work, we are currently working on Network Design and Dimensionning using Routing Models and taking into account network reliability.

REFERENCES

- [1] A. GIRARD, "Routing and Dimensionning in Circuit Switched Networks", Addison Wesley editors, to appear.
- [2] B. SANSO, F. SOUMIS and M. GENDREAU, "On the Evaluation of Telecommunications Network Reliability Using Routing Models" publication 623, Centre de recherche sur les transports, Université de Montréal, March 1989.
- [3] B. SANSO, "Fiabilité et routage dans un réseau de télécommunications", Ph.D. dissertation, publication 605, Centre de recherche sur les transports, Université de Montréal, October 1988.
- [4] K. KORTANEK and G. POLAK., "Network design and dynamic routing under queuing demand", Carnegie-Mellon University, working paper, 1984.
- [5] C. PALM, "Some observations on the Earlang formulae for busy-signal systems, TELE, 1, English edition, 1957.
- [6] V.O.K. LI and J.A. SILVERSTER, "Performance analysis of networks with unreliable components", IEEE Transactions on Communications, COM-32 (10), 1105-1110, October 1984.
- [7] R.F. GAEBLER and R.J. CHEN, "An efficient algorithm for enumerating states of a system with multimode unreliable components", working paper presented at the ORSA-TIMS St.Louis meeting, October 25-28, 1987.
- [8] J. GUÉLAT, "Algorithmes pour le problème d'affectation du trafic d'équilibre avec demandes fixes, comparaisons," publication 299, Centre de recherche sur les transports, Université de Montréal, May 1983.